

WHAT IS CLAIMED IS:

1. A method of identifying the location of exons within the genome of a species of organism comprising:

- 5 (a) contacting a sample comprising RNAs or nucleic acids derived therefrom from one or more cells of said species of organism with an array, said array comprising a positionally-addressable ordered array of polynucleotide probes bound to a solid support, said polynucleotide probes comprising a first plurality of at least 100 polynucleotide probes of different nucleotide sequences, each said different nucleotide sequence comprising a
10 sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, said contacting being under conditions conducive to hybridization between said RNAs or nucleic acids derived therefrom and said probes;
- 15 (b) identifying the one or more probes to which hybridization of one or more of said RNAs or nucleic acids derived therefrom occurs; and
- (c) identifying said genomic sequences for each said identified probe as the location of an exon within the genome of said species of organism.

20 2. The method of claim 1, wherein step (a) is repeated with RNAs or nucleic acids derived therefrom from a plurality of different cells of said species of organism.

3. The method of claim 1, wherein said array has in the range of 150 to 1,000 different polynucleotide probes per 1 cm².

25

4. The method of claim 1, wherein said array has in the range of 1,000 to 10,000 different polynucleotide probes per 1 cm².

5. The method of claim 1, wherein said array has in the range of 10,000 to
30 50,000 different polynucleotide probes per 1 cm².

6. The method of claim 1, wherein said array has greater than 50,000 different polynucleotide probes per 1 cm².

35

7. The method of claim 1, wherein the nucleotide sequences of the probes consist of up to 1000 nucleotides.

5 8. The method of claim 1, wherein the nucleotide sequences of the probes consist of in the range of 10-200 nucleotides.

9. The method of claim 1, wherein the nucleotide sequences of the probes consist of in the range of 80-120 nucleotides.

10 10. The method of claim 1, wherein the nucleotide sequences of the probes consist of in the range of 40-80 nucleotides.

11. The method of claim 1, wherein the nucleotide sequences of the probes consist of 60 nucleotides.

15 12. The method of claim 1, wherein said genomic sequences for different probes are overlapping in said genome.

13. The method of claim 1, wherein said genomic sequences for different probes
20 are overlapping in said genome from 10-50% of the length of each said different probe.

14. The method of claim 1, wherein said genomic sequences for different probes are adjacent in said genome.

25 15. The method of claim 1, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

16. The method as in one of claims 7-11, wherein said genomic sequences for different probes are overlapping in said genome.

30 17. The method as in one of claims 7-11, wherein said genomic sequences for different probes are overlapping in said genome from 10-50% of the length of each said different probe.

35

18. The method as in one of claims 7-11, wherein said genomic sequences for different probes are adjacent in said genome.

19. The method as in one of claims 7-11, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

20. The method of claim 1, wherein said organism is a eukaryote.

21. The method of claim 1, wherein said organism is a human.

22. The method of claim 1, wherein said organism is a plant.

23. The method of claim 1, wherein said organism is a mammal.

24. The method of claim 1, wherein said first plurality of polynucleotide probes is at least 1,000 probes.

25. The method of claim 1, wherein said first plurality of polynucleotide probes is at least 10,000 probes.

26. The method of claim 1, wherein said first plurality of polynucleotide probes is in the range of 1,000 to 50,000 probes.

27. The method of claim 1, wherein two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least 10 different genes.

28. The method of claim 1, wherein the distance between 5' ends of said sequential sites is always less than 500 bp, and wherein the genomic sequences for said first plurality of probes span a genomic region of at least 25,000 bp.

29. The method of claim 1, wherein two or more of said polynucleotide probes are complementary and hybridizable to sequences contained entirely within an intron, and wherein said ordered array does not comprise a second plurality of polynucleotide probes that do not comprise a sequence complementary and hybridizable to said genome of said

species of organism, said second plurality being of equal or greater number than said first plurality.

30. The method of claim 1, wherein two or more of said polynucleotide probes
5 are complementary and hybridizable to intron sequences of at least 10 different genes, and
wherein said ordered array does not comprise a second plurality of polynucleotide probes
that do not comprise a sequence complementary and hybridizable to said genome of said
species of organism, said second plurality being of equal or greater number than said first
plurality.

10

31. The method of claim 1, wherein:

(a) said polynucleotide probes further comprise a second plurality of
polynucleotide probes comprising a sequence complementary and hybridizable to said first
plurality; and

15

(b) said identifying step comprises using a hybridization signal generated in said
contacting step from said second plurality to filter a hybridization signal generated in said
contacting step from said first plurality.

32. The method of claim 1, wherein:

20

(a) said sample comprises RNAs or nucleic acids derived therefrom from (i) a
first cell or cells of a first tissue type or of a first condition, and (ii) a second cell or cells of
a second tissue type different from said first tissue type or of a second condition different
from said first condition; and

(b) said identifying step comprises comparing a hybridization signal generated in
25 said contacting step from said first cell or cells to a hybridization signal generated in said
contacting step from said second cell or cells.

33. The method of claim 1, wherein said plurality of probes is tiled across an
area predicted to contain, or known to contain, exons.

30

34. The method of claim 1, wherein said plurality of probes includes known
expressed sequence tags (ESTs) or predicted exons.

35. The method of claim 1, wherein each of said plurality of probes corresponds
35 to a predicted or known exon.

36. The method of claim 1, further comprising a sample comprising a population of cellular RNA or nucleic acid derived therefrom on the surface of said solid support such that said sample is in contact with said polynucleotide probes, under conditions conducive to hybridization between said population and said polynucleotide probes.

37. The method of claim 36 wherein said population is labeled.

38. The method of claim 36 wherein said population comprises total cellular mRNA or nucleic acid derived therefrom.

39. The method of claim 36 wherein said population comprises nucleic acids of at least 10,000 different sequences.

40. A method for identifying the approximate location of an intron-exon boundary in the genome of a species of organism comprising:

(a) contacting a sample comprising RNAs or nucleic acids derived therefrom from one or more cells of said species of organism with an array, said array comprising a positionally-addressable ordered array of polynucleotide probes bound to a solid support, said polynucleotide probes comprising a first plurality of at least 100 polynucleotide probes of different nucleotide sequences, each said different nucleotide sequence comprising a sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, said contacting being under conditions conducive to hybridization between said RNAs or nucleic acids derived therefrom and said probes;

(b) determining, for each probe in said first plurality, whether or not hybridization of one or more of said RNAs or nucleic acids derived therefrom occurs;

(c) identifying pairs of probes for which said respective genomic sequences are closest in the genome and wherein one probe of said pair is determined to hybridize in step (b) and the other probe of said pair is determined not to hybridize in step (b), wherein the part of the genome in between said respective sequences for said pair of probes is the approximate location of an intron-exon boundary in said genome.

41. The method of claim 40, wherein said array has in the range of 150 to 1,000 different polynucleotide probes per 1 cm².

42. The method of claim 40, wherein said array has in the range of 1,000 to 10,000 different polynucleotide probes per 1 cm².

43. The method of claim 40, wherein said array has in the range of 10,000 to 50,000 different polynucleotide probes per 1 cm².

44. The method of claim 40, wherein said array has greater than 50,000 different polynucleotide probes per 1 cm².

45. The method of claim 40, wherein the nucleotide sequences of the probes consist of no more than 1,000 nucleotides.

46. The method of claim 40, wherein the nucleotide sequences of the probes consist of in the range of 10-200 nucleotides.

47. The method of claim 40, wherein the nucleotide sequences of the probes consist of in the range of 10-40 nucleotides.

48. The method of claim 40, wherein the nucleotide sequences of the probes consist of in the range of 40-80 nucleotides.

49. The method of claim 40, wherein the nucleotide sequences of the probes consist of 60 nucleotides.

50. The method of claim 40, wherein said genomic sequences for different probes are overlapping in said genome.

51. The method of claim 40, wherein said genomic sequences for different probes are overlapping in said genome from 50-90% of the length of each said different probe.

52. The method of claim 40, wherein said genomic sequences for different probes are overlapping in said genome from 70-80% of the length of each said different probe.

53. The method of claim 40, wherein said genomic sequences for different probes are overlapping at all but one base pair.

54. The method of claim 40, wherein said genomic sequences for different probes are adjacent in said genome.

55. The method as in one of claims 45-49, wherein said genomic sequences for different probes are overlapping in said genome.

56. The method as in one of claims 45-49, wherein said genomic sequences for different probes are overlapping in said genome from 50-90% of the length of each said different probe.

57. The method as in one of claims 45-49, wherein said genomic sequences for different probes are overlapping in said genome from 70-80% of the length of each said different probe.

58. The method as in one of claims 45-49, wherein said genomic sequences for different probes are overlapping at all but one base pair.

59. The method as in one of claims 45-49, wherein said genomic sequences for different probes are adjacent in said genome.

60. The method of claim 40, wherein said organism is a eukaryote.

61. The method of claim 40, wherein said organism is a human.

62. The method of claim 40, wherein said organism is a plant.

63. The method of claim 40, wherein said organism is a mammal.

64. The method of claim 40, wherein said first plurality of polynucleotide probes is at least 1,000 probes.

65. The method of claim 40, wherein said first plurality of polynucleotide probes is at least 10,000 probes.

5 66. The method of claim 40, wherein said first plurality of polynucleotide probes is in the range of 1,000 to 50,000 probes.

67. The method of claim 40, wherein two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least 10 different genes.

10 68. The method of claim 40, wherein the distance between 5' ends of said sequential sites is always less than 500 bp, wherein the genomic sequences for said first plurality of probes span a genomic region of at least 25,000 bp.

15 69. The method of claim 40, wherein two or more of said polynucleotide probes are complementary and hybridizable to sequences contained entirely within an intron, and wherein said ordered array does not comprise a second plurality of polynucleotide probes that do not comprise a sequence complementary and hybridizable to said genome of said species of organism, said second plurality being of equal or greater number than said first plurality.

20

70. The method of claim 40, wherein two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least 10 different genes, and wherein said ordered array does not comprise a second plurality of polynucleotide probes that do not comprise a sequence complementary and hybridizable to said genome of said species of organism, said second plurality being of equal or greater number than said first plurality.

71. The method of claim 40, further comprising a sample comprising a population of cellular RNA or nucleic acid derived therefrom on the surface of said solid support such that said sample is in contact with said polynucleotide probes, under conditions conducive to hybridization between said population and said polynucleotide probes.

35 72. The method of claim 71 wherein said population is labeled.

73. The method of claim 71 wherein said population comprises total cellular mRNA or nucleic acid derived therefrom.

74. The method of claim 71 wherein said population comprises nucleic acids of
5 at least 10,000 different sequences.

75. A method of determining the amino-terminus of a protein, comprising:
(a) contacting a sample comprising RNAs or nucleic acids derived therefrom
from one or more cells of said species of organism with an array, said array comprising a
10 positionally-addressable ordered array of polynucleotide probes bound to a solid support,
said polynucleotide probes comprising a first plurality of at least 100 polynucleotide probes
of different nucleotide sequences, each said different nucleotide sequence comprising a
sequence complementary and hybridizable to a different genomic sequence of the same
species of organism, said respective genomic sequences for the probes being found at
15 sequential sites in said genome of said species of organism, said contacting being under
conditions conducive to hybridization between said RNAs or nucleic acids derived
therefrom and said probes;

(b) identifying a probe to which hybridization of one or more of said RNAs or
nucleic acids derived therefrom occurs, and for which said genomic sequence is within a
20 genomic region predicted to encode a 5' untranslated region of a mRNA; and

(c) identifying a start codon in said genome; said start codon being the nearest
start codon 3' to said genomic sequence for said identified probe, wherein an internal
ribosome entry site appears encoded in the genome 5' to said start codon and within said 5'
untranslated region, and wherein the amino-terminus of said protein is encoded by the
25 sequence of the genome immediately 3' to said start codon.

76. The method of claim 75, wherein said array has in the range of 150 to 1,000
different polynucleotide probes per 1 cm².

30 77. The method of claim 75, wherein said array has in the range of 1,000 to
10,000 different polynucleotide probes per 1 cm².

78. The method of claim 75, wherein said array has in the range of 10,000 to
50,000 different polynucleotide probes per 1 cm².

35

79. The method of claim 75, wherein said array has greater than 50,000 different polynucleotide probes per 1 cm².

5 80. The method of claim 75, wherein the nucleotide sequences of the probes consist of no more than 1,000 nucleotides.

81. The method of claim 75, wherein the nucleotide sequences of the probes consist of in the range of 10-200 nucleotides.

10 82. The method of claim 75, wherein the nucleotide sequences of the probes consist of in the range of 10-40 nucleotides.

83. The method of claim 75, wherein the nucleotide sequences of the probes consist of in the range of 40-80 nucleotides.
15

84. The method of claim 75, wherein the nucleotide sequences of the probes consist of in the range of 80-120 nucleotides.

85. The method of claim 75, wherein the nucleotide sequences of the probes
20 consist of 60 nucleotides.

86. The method of claim 75, wherein said genomic sequences for different probes are overlapping in said genome.

25 87. The method of claim 75, wherein said genomic sequences for different probes are overlapping in said genome from 10-50% of the length of each said different probe.

88. The method of claim 75, wherein said genomic sequences for different
30 probes are overlapping in said genome from 50-90% of the length of each said different probe.

89. The method of claim 75, wherein said genomic sequences for different
35 probes are overlapping in said genome from 70-80% of the length of each said different probe.

90. The method of claim 75, wherein said genomic sequences for different probes are overlapping at all but one base pair.

91. The method of claim 75, wherein said genomic sequences for different probes are adjacent in said genome.

92. The method as in one of claims 80-85, wherein said genomic sequences for different probes are overlapping in said genome.

93. The method as in one of claims 80-85, wherein said genomic sequences for different probes are overlapping in said genome from 10-50% of the length of each said different probe.

94. The method as in one of claims 80-85, wherein said genomic sequences for different probes are overlapping in said genome from 50-90% of the length of each said different probe.

95. The method as in one of claims 80-85, wherein said genomic sequences for different probes are overlapping in said genome from 70-80% of the length of each said different probe.

96. The method as in one of claims 80-85, wherein said genomic sequences for different probes are overlapping at all but one base pair.

97. The method as in one of claims 80-85, wherein said genomic sequences for different probes are adjacent in said genome.

98. The method of claim 75, wherein said organism is a eukaryote.

99. The method of claim 75, wherein said organism is a human.

100. The method of claim 75, wherein said organism is a plant.

101. The method of claim 75, wherein said organism is a mammal.

102. The method of claim 75, wherein said first plurality of polynucleotide probes is at least 1,000 probes.

103. The method of claim 75, wherein said first plurality of polynucleotide probes is at least 10,000 probes.

104. The method of claim 75, wherein said first plurality of polynucleotide probes is in the range of 1,000 to 50,000 probes.

105. The method of claim 75, wherein two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least 10 different genes.

106. The method of claim 75; wherein the distance between 5' ends of said sequential sites is always less than 500 bp, wherein the genomic sequences for said first plurality of probes span a genomic region of at least 25,000 bp.

107. The method of claim 75, wherein two or more of said polynucleotide probes are complementary and hybridizable to sequences contained entirely within an intron, and wherein said ordered array does not comprise a second plurality of polynucleotide probes that do not comprise a sequence complementary and hybridizable to said genome of said species of organism, said second plurality being of equal or greater number than said first plurality.

108. The method of claim 75, wherein two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least 10 different genes, and wherein said ordered array does not comprise a second plurality of polynucleotide probes that do not comprise a sequence complementary and hybridizable to said genome of said species of organism, said second plurality being of equal or greater number than said first plurality.

109. The method of claim 75, further comprising a sample comprising a population of cellular RNA or nucleic acid derived therefrom on the surface of said solid support such that said sample is in contact with said polynucleotide probes, under conditions conducive to hybridization between said population and said polynucleotide probes.

110. The method of claim 109 wherein said population is labeled.

111. The method of claim 109 wherein said population comprises total cellular mRNA or nucleic acid derived therefrom.

5

112. The method of claim 109 wherein said population comprises nucleic acids of at least 10,000 different sequences.

113. A method of determining the probability that an individual nucleotide within
10 the genome of a species of organism is expressed comprising:

(a) contacting a sample comprising RNAs or nucleic acids derived therefrom from one or more cells of said species of organism with an array, said array comprising a positionally-addressable ordered array of polynucleotide probes bound to a solid support, said polynucleotide probes comprising a plurality of at least 100 polynucleotide probes of
15 different nucleotide sequences, each said different nucleotide sequence comprising a sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, said contacting being under conditions conducive to hybridization between said RNAs or nucleic acids derived
20 therefrom and said probes;

(b) determining, for each probe in said plurality, whether or not hybridization of one or more of said RNAs or nucleic acids derived therefrom occurs;

(c) based on the determinations in step (b), calculating the probability that a given nucleotide within said respective genomic sequences is expressed.

25

114. The method of claim 113 wherein said calculating further comprises considering EST data.

115. A computer-implemented method for designing probes for an array
30 comprising:

(a) inputting a genomic sequence of a species of organism into a computer;

(b) analyzing said genomic sequence to exclude repetitive elements, simple repeats, or polyX repeats;

35

(c) analyzing said genomic sequence not excluded in step (b) to generate a list of sequences of a selected length, said sequences being complementary to sequential sites in said genomic sequence; and

(d) outputting said list of sequences.

5

116. The computer-implemented method of claim 115, further comprising:

(e) applying a tiling strategy to said list of sequences to generate a list of tiled sequences; and

(f) outputting said list of tiled sequences.

10

117. A computer system for identifying the location of exons within the genome of a species of organism, said computer system comprising:

one or more processor units; and

one or more memory units connected to said one or more processor units, said one or more memory units containing one or more programs which cause said one or more processor units to execute steps of:

(a) receiving a first data structure comprising a first plurality of measured hybridization signals from an array comprising a positionally-addressable ordered array of polynucleotide probes bound to a solid support, said polynucleotide probes comprising a second plurality of at least 100 polynucleotide probes of different nucleotide sequences, each said different nucleotide sequence comprising a sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, said array contacting a sample comprising RNAs or nucleic acids derived therefrom from one or more cells of said species of organism with said array, said contacting being under conditions conducive to hybridization between said RNAs or nucleic acids derived therefrom and said probes;

(b) receiving a second data structure comprising the nucleotide sequence of said genome of said organism;

(c) receiving a third data structure comprising the nucleotide sequence of said second plurality of polynucleotide probes, said third data structure identifying the positional location of each said probe on said array;

(d) identifying the one or more probes to which hybridization of one or more of said RNAs or nucleic acids derived therefrom occurs;

35

- (e) identifying said genomic sequences for each said identified probe as the location of an exon within the genome of said species of organism; and
- (f) outputting the locations of said exons with respect to the nucleotide sequence of said genome of said organism.

5

118. A computer system for identifying an intron-exon junction boundary in the genome of a species or organism, said computer system comprising:

one or more processor units; and

one or more memory units connected to said one or more processor units, said one
10 or more memory units containing one or more programs which cause said one or more processor units to execute steps of:

(a) receiving a first data structure comprising a first plurality of measured hybridization signals from an array comprising a positionally-addressable ordered array of polynucleotide probes bound to a solid support, said polynucleotide probes comprising a
15 second plurality of at least 100 polynucleotide probes of different nucleotide sequences, each said different nucleotide sequence comprising a sequence complementary and hybridizable to a different genomic sequence comprising a sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said
20 respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, said array contacting a sample comprising RNAs or nucleic acids derived therefrom from one or more cells of said species of organism with said array, said contacting being under conditions conducive to hybridization between said RNAs or nucleic acids derived therefrom and said probes;

(b) receiving a second data structure comprising the nucleotide sequence of said
25 genome of said organism;

(c) receiving a third data structure comprising the nucleotide sequence of said second plurality of polynucleotide probes, said third data structure identifying the positional location of each said probe on said array;

(d) determining, for each probe in said second plurality, whether or not
30 hybridization of one or more of said RNAs or nucleic acids derived therefrom occurs;

(e) identifying pairs of probes for which said respective genomic sequences are closest in the genome and wherein one probe of said pair is determined to hybridize in step (d) and the other probe of said pair is determined not to hybridize in step (d), wherein the part of the genome in between said respective sequences for said pair of probes is the
35 location of an intron-exon boundary in said genome; and

(f) outputting the locations of said intron-exon boundaries with respect to the nucleotide sequence of said genome of said organism.

119. A computer program product for identifying the location of exons within the
5 genome of a species of organism, the computer program product for use in conjunction with
a computer having a memory and a processor, the computer program product comprising a
computer readable storage medium having a computer program mechanism encoded
thereon, wherein said computer program mechanism can be loaded into the one or more
memory units of a computer and cause the one or more processor units of the computer to
10 execute the steps of:

(a) receiving a first data structure comprising a first plurality of measured
hybridization signals from an array comprising a positionally-addressable ordered array of
polynucleotide probes bound to a solid support, said polynucleotide probes comprising a
second plurality of at least 100 polynucleotide probes of different nucleotide sequences,
15 each said different nucleotide sequence comprising a sequence complementary and
hybridizable to a different genomic sequence of the same species of organism, said
respective genomic sequences for the probes being found at sequential sites in said genome
of said species of organism, said array contacting a sample comprising RNAs or nucleic
acids derived therefrom from one or more cells of said species of organism with said array,
20 said contacting being under conditions conducive to hybridization between said RNAs or
nucleic acids derived therefrom and said probes;

(b) receiving a second data structure comprising the nucleotide sequence of said
genome of said organism;

(c) receiving a third data structure comprising the nucleotide sequence of said
25 second plurality of polynucleotide probes, said third data structure identifying the positional
location of each said probe on said array;

(d) identifying the one or more probes to which hybridization of one or more of
said RNAs or nucleic acids derived therefrom occurs;

(e) identifying said genomic sequences for each said identified probe as the
30 location of an exon within the genome of said species of organism; and

(f) outputting the locations of exon boundaries with respect to the nucleotide
sequence of said genome of said organism.

120. A computer program product for identifying an intron-exon junction
35 boundary in the genome of a species of organism, the computer program product for use in

conjunction with a computer having a memory and a processor, the computer program product comprising a computer readable storage medium having a computer program mechanism encoded thereon, wherein said computer program mechanism can be loaded into the one or more memory units of a computer and cause the one or more processor units of
5 the computer to execute the steps of:

(a) receiving a first data structure comprising a first plurality of measured hybridization signals from an array comprising a positionally-addressable ordered array of polynucleotide probes bound to a solid support, said polynucleotide probes comprising a second plurality of at least 100 polynucleotide probes of different nucleotide sequences,
10 each said different nucleotide sequence comprising a sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, said array contacting a sample comprising RNAs or nucleic acids derived therefrom from one or more cells of said species of organism with said array,
15 said contacting being under conditions conducive to hybridization between said RNAs or nucleic acids derived therefrom and said probes;

(b) receiving a second data structure comprising the nucleotide sequence of said genome of said organism;

(c) receiving a third data structure comprising the nucleotide sequence of said second plurality of polynucleotide probes, said third data structure identifying the positional location of each said probe on said array;

(d) determining, for each probe in said second plurality, whether or not hybridization of one or more of said RNAs or nucleic acids derived therefrom occurs;

(e) identifying pairs of probes for which said respective genomic sequences are
25 closest in the genome and wherein one probe of said pair is determined to hybridize in step (d) and the other probe of said pair is determined not to hybridize in step (d), wherein the part of the genome in between said respective sequences for said pair of probes is the location of an intron-exon boundary in said genome; and

(f) outputting the locations of said intron-exon boundaries with respect to the
30 nucleotide sequence of said genome of said organism.

121. An array, comprising:
a positionally-addressable ordered array of polynucleotide probes bound to a solid support; and

5 said polynucleotide probes comprising a first plurality of at least 100 polynucleotide probes of different nucleotide sequences, each said different nucleotide sequence comprising a sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, wherein two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least
10 10 different genes.

122. An array, comprising:
a positionally-addressable ordered array of polynucleotide probes bound to a solid support; and

15 said polynucleotide probes comprising a first plurality of at least 100 polynucleotide probes of different nucleotide sequences, each said different nucleotide sequence comprising a sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, wherein the distance between
20 5' ends of said sequential sites is always less than 500 bp, wherein the genomic sequences for said first plurality of probes span a genomic region of at least 25,000 bp.

123. An array, comprising:
a positionally-addressable ordered array of polynucleotide probes bound to a solid
25 support;

said polynucleotide probes comprising a first plurality of at least 100 polynucleotide probes of different nucleotide sequences, each said different nucleotide sequence comprising a sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said respective genomic sequences for the probes being found
30 at sequential sites in said genome of said species of organism;

wherein two or more of said polynucleotide probes are complementary and hybridizable to sequences contained entirely within an intron; and

wherein said ordered array does not comprise a second plurality of polynucleotide probes that do not comprise a sequence complementary and hybridizable to said genome of
35

said species of organism, said second plurality being of equal or greater number than said first plurality.

124. An array, comprising:
- 5 a positionally-addressable ordered array of polynucleotide probes bound to a solid support;
- said polynucleotide probes comprising a first plurality of at least 100 polynucleotide probes of different nucleotide sequences, each said different nucleotide sequence comprising a sequence complementary and hybridizable to a different genomic sequence of
- 10 the same species of organism, said respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, wherein two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least 10 different genes; and
- wherein said ordered array does not comprise a second plurality of polynucleotide
- 15 probes that do not comprise a sequence complementary and hybridizable to said genome of said species of organism, said second plurality being of equal or greater number than said first plurality.

125. An array as in one of claims 121-124, wherein the array has in the range of
- 20 150 to 1,000 different polynucleotide probes per 1 cm².

126. An array as in one of claims 121-124, wherein the array has in the range of 1,000 to 10,000 different polynucleotide probes per 1 cm².

- 25 127. An array as in one of claims 121-124, wherein the array has in the range of 10,000 to 50,000 different polynucleotide probes per 1 cm².

128. An array as in one of claims 121-124, wherein the array has greater than 50,000 different polynucleotide probes per 1 cm².

- 30 129. An array as in one of claims 121-124, wherein said genomic sequences for different probes are overlapping in said genome.

130. An array as in one of claims 121-124, wherein said genomic sequences for
- 35 different probes are adjacent in said genome.

131. An array as in one of claims 121-124, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

5 132. An array as in one of claims 121-124, wherein the nucleotide sequences of the probes consist of no more than 1,000 nucleotides.

133. An array as in claim 132, wherein said genomic sequences for different probes are overlapping in said genome.

10 134. An array as in claim 132, wherein said genomic sequences for different probes are adjacent in said genome.

135. An array as in claim 132, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

15 136. An array as in one of claims 121-124, wherein the nucleotide sequences of the probes consist of in the range of 10-200 nucleotides.

20 137. An array as in claim 136, wherein said genomic sequences for different probes are overlapping in said genome.

138. An array as in claim 136, wherein said genomic sequences for different probes are adjacent in said genome.

25 139. An array as in claim 136, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

30 140. An array as in one of claims 121-124, wherein the nucleotide sequences of the probes consist of in the range of 10-30 nucleotides.

141. An array as in claim 140, wherein said genomic sequences for different probes are overlapping in said genome.

35 142. An array as in claim 140, wherein said genomic sequences for different probes are adjacent in said genome.

143. An array as in claim 140, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

144. An array as in one of claims 121-124, wherein the nucleotide sequences of
5 the probes consist of in the range of 20-50 nucleotides.

145. An array as in claim 144, wherein said genomic sequences for different probes are overlapping in said genome.

10 146. An array as in claim 144, wherein said genomic sequences for different probes are adjacent in said genome.

147. An array as in claim 144, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

15

148. An array as in one of claims 121-124, wherein the nucleotide sequences of the probes consist of in the range of 40-80 nucleotides.

149. An array as in claim 148, wherein said genomic sequences for different
20 probes are overlapping in said genome.

150. An array as in claim 148, wherein said genomic sequences for different probes are adjacent in said genome.

25 151. An array as in claim 148, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

152. An array as in one of claims 121-124, wherein the nucleotide sequences of the probes consist of in the range of 50-150 nucleotides.

30

153. An array as in claim 152, wherein said genomic sequences for different probes are overlapping in said genome.

154. An array as in claim 152, wherein said genomic sequences for different
35 probes are adjacent in said genome.

155. An array as in claim 152, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

5 156. An array as in one of claims 121-124, wherein the nucleotide sequences of the probes consist of 60 nucleotides.

157. An array as in claim 156, wherein said genomic sequences for different probes are overlapping in said genome.

10 158. An array as in claim 156, wherein said genomic sequences for different probes are adjacent in said genome.

159. An array as in claim 156, wherein said genomic sequence for each probe is spaced apart from that for other probes in said genome by less than 200 bp.

15

160. An array as in one of claims 121-124, wherein said organism is a eukaryote.

161. An array as in one of claims 121-124, wherein said organism is a human.

20 162. An array as in one of claims 121-124, wherein said organism is a plant.

163. An array as in one of claims 121-124, wherein said organism is a mammal.

25 164. An array as in one of claims 121-124, wherein said first plurality of polynucleotide probes is at least 1,000 probes.

165. An array as in one of claims 121-124, wherein said first plurality of polynucleotide probes is at least 10,000 probes.

30 166. An array as in one of claims 121-124, wherein said first plurality of polynucleotide probes is in the range of 1,000 to 50,000 probes.

35 167. An array as in one of claims 121-124, wherein said polynucleotide probes comprising sequences corresponding to repetitive elements, simple repeats, or polyX repeats are excluded as probes.

168. An array as in one of claims 121-124, further comprising a sample comprising a population of cellular RNA or nucleic acid derived therefrom on the surface of said solid support such that said sample is in contact with said polynucleotide probes, under conditions conducive to hybridization between said population and said polynucleotide probes.

169. An array as in claim 168 wherein said population is labeled.

170. An array as in claim 168 wherein said population comprises total cellular mRNA or nucleic acid derived therefrom.

171. An array as in claim 168 wherein said population comprises nucleic acids of at least 10,000 different sequences.

172. An array as in claim 121 or 124 wherein said at least 10 different genes comprises at least 20 different genes.

173. An array as in claim 121 or 124 wherein said at least 10 different genes comprises at least 50 different genes.

174. An array as in claim 121 or 124 wherein said at least 10 different genes comprises at least 200 different genes.

175. An array as in claim 121 or 124 wherein said at least 10 different genes comprises at least 1,000 different genes.

176. An array as in claim 123 or 124, wherein said ordered array does not comprise one or more matched probes for each of said polynucleotide probes in said first plurality, the sequence of said matched probes varying only in the identity of a single nucleotide at the same position relative to said polynucleotide probe.

177. A method for preparing an array comprising synthesizing a plurality of polynucleotide probes on a solid support, wherein:

said polynucleotide probes are ordered on said solid support so as to be positionally addressable;

said polynucleotide probes comprise a plurality of at least 100 polynucleotide probes of different nucleotide sequences;

each said different nucleotide sequence comprises a sequence complementary and hybridizable to a different genomic sequence of the same species of organism;

5 said respective genomic sequences for said polynucleotide probes are found at sequential sites in the said genome of said species of organism; and

two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least 10 different genes.

10 178. A method for preparing an array comprising synthesizing a plurality of polynucleotide probes on a solid support, wherein:

said polynucleotide probes are ordered on said solid support so as to be positionally-addressable;

15 said polynucleotide probes comprise a plurality of at least 100 polynucleotide probes of different nucleotide sequences;

each said different nucleotide sequence comprises a sequence complementary and hybridizable to a different genomic sequence of the same species of organism;

said respective genomic sequences for said polynucleotide probes are found at sequential sites in the said genome of said species of organism;

20 the distance between 5' ends of said sequential sites is always less than 500 bp; and the genomic sequences for said plurality of probes span a genomic region of at least 25,000 bp.

25 179. A method for preparing an array comprising synthesizing a plurality of polynucleotide probes on a solid support, wherein:

said polynucleotide probes are ordered on said solid support so as to be positionally-addressable;

said polynucleotide probes comprise a first plurality of at least 100 polynucleotide probes of different nucleotide sequences;

30 each said different nucleotide sequence comprises a sequence complementary and hybridizable to a different genomic sequence of the same species of organism;

said respective genomic sequences for said polynucleotide probes are found at sequential sites in the said genome of said species of organism;

35 two or more of said polynucleotide probes are complementary and hybridizable to sequences contained entirely within an intron; and

said ordered array does not comprise a second plurality of polynucleotide probes that do not comprise a sequence complementary and hybridizable to said genome of said species of organism, said second plurality being of equal or greater number than said first plurality.

5 180. A method for preparing an array comprising synthesizing a plurality of polynucleotide probes on a solid support, wherein:

said polynucleotide probes are ordered on said solid support so as to be positionally-addressable;

10 said polynucleotide probes comprise a first plurality of at least 100 polynucleotide probes of different nucleotide sequences;

each said different nucleotide sequence comprises a sequence complementary and hybridizable to a different genomic sequence of the same species of organism;

said respective genomic sequences for said polynucleotide probes are found at sequential sites in the said genome of said species of organism;

15 two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least 10 different genes; and

said ordered array does not comprise a second plurality of polynucleotide probes that do not comprise a sequence complementary and hybridizable to said genome of said species of organism, said second plurality being of equal or greater number than said first plurality.

20

181. A method for preparing an array comprising placing a plurality of polynucleotide probes on a solid support, wherein:

said polynucleotide probes are ordered on said solid support so as to be positionally-addressable;

25 said polynucleotide probes comprise a plurality of at least 100 polynucleotide probes of different nucleotide sequences;

each said different nucleotide sequence comprises a sequence complementary and hybridizable to a different genomic sequence of the same species of organism;

30 said respective genomic sequences for said polynucleotide probes are found at sequential sites in the said genome of said species of organism; and

two or more of said polynucleotide probes are complementary and hybridizable to intron sequences of at least 10 different genes.

35

182. A method of determining whether respective sequences encoded by two or more exons are indicated to be present in a single mRNA transcript, comprising:

5 (a) contacting, under conditions conducive to hybridization, a plurality of samples with a positionally-addressable array containing probes that are identified as complementary and hybridizable to potential exons in the genome of a species of organism, said samples each comprising RNAs or nucleic acids derived therefrom from a cell of said species of organism exposed to a different condition; and

10 (b) determining whether the level of hybridization to one or more probes complementary and hybridizable to RNA or nucleic acids derived therefrom encoded by a first potential exon and the level of hybridization of one or more probes complementary and hybridizable to RNA or nucleic acids derived therefrom encoded by a potential neighboring exon are correlated over the plurality of samples, wherein if said levels are correlated, said respective sequences encoded by said first potential exon and said neighboring exon are indicated to be present in a single RNA transcript.

15

183. The method of claim 182, further comprising:

(c) determining whether the level of hybridization to one or more probes complementary and hybridizable to RNA or nucleic acids derived therefrom encoded by an exon additional to said first exon and said neighboring exon, and the respective levels of
20 hybridization of one or more probes complementary and hybridizable to RNA or nucleic acids derived therefrom encoded by said first exon and said neighboring exon, are correlated over a plurality of samples, wherein if said levels are correlated, said respective sequences encoded by said first potential exon, said neighboring exon and said additional exon are indicated to be present in said single RNA transcript; and

25 (d) repeating step (c) until no further exons are indicated to be present in said single RNA transcript.

184. A method of determining the probability that an individual nucleotide within the genome of a species of organism is expressed in response to a condition, comprising:

30 (a) contacting a first sample and a second sample, both comprising RNAs or nucleic acids derived therefrom from one or more cells of said species of organism, with an array under conditions conducive to hybridization, said array comprising a positionally-addressable ordered array of polynucleotide probes bound to a solid support, said polynucleotide probes comprising a plurality of at least 100 polynucleotide probes of
35 different nucleotide sequences, each said different nucleotide sequence comprising a

sequence complementary and hybridizable to a different genomic sequence of the same species of organism, said respective genomic sequences for the probes being found at sequential sites in said genome of said species of organism, said contacting being under conditions conducive to hybridization between said RNAs or nucleic acids derived

5 therefrom and said probes, said first sample being derived from cells exposed to a condition, said second sample being derived from cells not exposed to said condition, said RNAs or nucleic acids derived therefrom from said first and second samples being differentially labeled;

(b) determining, for each probe in said plurality, whether or not
10 hybridization of one or more of said RNAs or nucleic acids derived therefrom from said first or second samples occurs; and

(c) based on the determinations in step (b), calculating the probability that a given nucleotide within said respective genomic sequences is expressed in response to a condition.

15

20

25

30

35